FINITE SEMIFIELDS AND PROJECTIVE PLANES

Thesis by

Donald Ervin Knuth

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1963

ACKNOWLEDGMENTS

I am indebted to Professor Marshall Hall, Jr., for his inspiration and guidance as I pursued this work. Thanks are also due to California Institute of Technology for providing scholarships to support my studies.

Many of the results of this paper were suggested and/or confirmed by calculations on a Burroughs 220 digital computer, and I am grateful to the Burroughs Corporation for providing computer time.

ABSTRACT

This paper makes contributions to the structure theory of finite semifields, i.e., of finite nonassociative division algebras with unit. It is shown that a semifield may be conveniently represented as a 3-dimensional array of numbers, and that matrix multiplications applied to each of the three dimensions correspond to the concept of isotopy. The six permutations of three coordinates yield a new way to obtain projective planes from a given plane. Several new classes of semifields are constructed; in particular one class, called the binary semifields, provides an affirmative answer to the conjecture that there exist non-Desarguesian projective planes of all orders 2^n, if n is greater than 3. With the advent of binary semifields, the gap between necessary and sufficient conditions on the possible orders of semifields has disappeared.

TABLE OF CONTENTS

I. INTRODUCTION

In this paper, the term semifield is used to describe an algebraic system which satisfies all properties of a field except for the commutativity and associativity of multiplication. Semifields are of special interest today because the projective planes constructed from them have rather remarkable properties.

It is easy to show that the order of a proper semifield, i.e., a finite semifield which is not a field, must be p^n, where p is a prime, n is an integer greater than 2, and p^n is greater than 8. But the question whether such systems exist for all p^n meeting these requirements has been in doubt for many years.

Various constructions of proper semifields have appeared in the literature; the earliest works in this area gave many constructions for odd primes p and eventually the existence of proper semifields for all possible orders p^n was known, as long as p was odd. Constructions appeared later for the case 2^n, but in every case the exponent n was composite. Therefore, the existence of proper semifields of order 2^q, where q is a prime number greater than 3, was still in doubt. In fact, no projective planes of these orders 2^q were known, except for the Desarguesian planes derived from a field. But this question is settled in Section VIII of the present paper, where a class of proper semifields including all of the missing orders is exhibited.

This paper is essentially self-contained; the theory of semifields is developed from the beginning, requiring no previous familiarity with the subject on the part of the reader. Section II contains a review of the basic definitions and theory of semifields, along with illustrations of some interesting semifields of order 16.

In Section III, we describe a "homogeneous" notation for representing points and lines of an arbitrary projective plane in terms of its ternary ring. The concept of isotopy is generalized to apply to arbitrary ternary rings, and a simple method for mechanically constructing all ternary rings isotopic to a given one is presented. Some known theorems about collineations of planes, and of semifield planes in particular, are proved concisely using the homogeneous notation. Finally, the question of whether a nonlinear "isotopy" can yield new semifields is considered.

Cubical arrays of numbers, of arbitrary finite dimension, are the subject of Section IV. First, operations of transposition and multiplication are discussed. Then the notion of a nonsingular hypercube is introduced. Semifields are shown to be equivalent to a certain type of 3-dimensional cubical array, and the projective planes coordinatized by semifields are in 1-1 correspondence with equivalence classes of nonsingular 3-cubes.

By transposition of a 3-cube, up to five new projective planes can be constructed from a single semifield plane. This construction is the topic of Section V. By exhibiting all of the semifield planes of order 32, with their interrelations and collineation groups, we give examples of transposed planes.

The finite semifields which have appeared in the literature are surveyed in Section VI, including a discussion of all semifields of order 16; of some commutative semifields due to Dickson; of the twisted fields due to Albert; and of seminuclear extensions due to Sandler.

A general class of quadratic extensions is considered in Section VII, in which the semifield is a vector space of dimension 2 over a so-called weak nucleus F. In particular, all quadratic extensions of F, for which F is equal to any two of the nuclei, are constructed, thus generalizing a result due to Hughes and Kleinfeld.

The last section deals with binary semifields, a new class of semifields of orders 2^n, where $n \neq 3$ and $n \neq 2^m$. It is shown that many of these systems of the same order are isotopic. Some of these systems having n automorphisms are exhibited, and the autotopism groups are partially determined.

The symbol "∎" will be used throughout this paper to mean: "This completes the proof of the theorem," or "This is all the proof of the theorem which will be given here."

The most interesting results of this paper are found perhaps in theorems 3.3.1, 3.3.2, 3.3.4, 4.4.2, 4.5.2, 5.1.1, 5.2.1, 7.2.1, 7.4.1, 8.2.1, 8.4.2, and 8.5.1.

II. SEMIFIELDS AND PRE-SEMIFIELDS.

We are concerned with a certain type of algebraic system, called a semifield. Such a system has several names in the literature, where it is called, for example, a "nonassociative division ring" or a "distributive quasifield." Since these terms are somewhat lengthy, and since we make frequent reference to such systems in this paper, the more convenient name semifield will be used.

2.1. Definition of semifield. A finite semifield S is a finite algebraic system containing at least two elements; S possesses two binary operations, addition and multiplication, which satisfy the following axioms.

A1. Addition is a group, with identity element 0.

A2. If $ab = 0$, then either $a = 0$ or $b = 0$.

A3. $a(b + c) = ab + ac$; $(a + b)c = ac + bc$.

A4. There is an element 1 in S such that $1a = a1 = a$.

Throughout this paper, the term semifield will always be used to denote a finite semifield. The definition given here would actually be insufficient to define infinite semifields, for the stronger condition that the equations $ax = b$ and $ya = b$ are uniquely solvable for x,y would necessarily replace axiom A2. Notice that a semifield is much like a field, except that multiplication of nonzero elements is required to be merely a loop instead of a group.

Every field is a semifield; the term proper semifield will mean a semifield which is not a field; i.e., there exist elements a,b,c such that $(ab)c \neq a(bc)$ in a proper semifield.

2.2. Examples. The following remarkable system V is a proper semifield with 16 elements: Let F be the field $GF(4)$, so that F has the elements $0, 1, \omega, \omega^2 = 1+\omega$. The elements of V are of the form $u + \lambda v$, where $u,v \in F$. Addition is defined in an obvious way:

$$(u + \lambda v) + (x + \lambda y) = (u+v) + \lambda(x+y) \tag{2.1}$$

using the addition of F. Multiplication is also defined in terms of the multiplication and addition of F, using the following rule:

$$(u + \lambda v)(x + \lambda y) = (ux + v^2 y) + \lambda(vx + u^2 y + v^2 y^2). \tag{2.2}$$

It is clear that F is embedded in V, and also that A1, A3, and A4 hold. To demonstrate A2, suppose that $(u + \lambda v)(x + \lambda v) = 0$. Then, in particular, $ux + v^2 y = 0$, so that, if neither of the original factors is zero, there is a nonzero element $z \in F$, such that

$$x = zv^2, \; y = zu.$$

Then $$vx + u^2 y + v^2 y^2 = zv^3 + zu^3 + z^2 u^2 v^2 = 0.$$

But this is impossible in F, unless $u = v = 0$.

V is certainly a proper semifield, since V is not commutative. But there is still a good deal of associativity present in V; in fact $(ab)c = a(bc)$ if any two of a,b,c are in F.

The remarkable property of V is that it possesses 6 automorphisms, while the field of 16 elements possesses only 4. The automorphisms σ_{ij} are given by:

$$(u + \lambda v)\sigma_{ij} = u^j + \lambda\omega^i v^j \quad \text{for } i = 0,1,2;\ j = 1,2.$$

(No other semifield of order 16 has as many automorphisms.) V is also anti-isomorphic to itself, under the mapping $(u + \lambda v)\tau = u + \lambda v^2$.

Other examples of semifields which have properties in common with V will be discussed later. We will remark here, however, that if we had defined multiplication by the rule

$$(u + \lambda v)(x + \lambda y) = (ux + \omega v^2 y) + \lambda(vx + u^2 y) \tag{2.3}$$

rather than as in (2.1), we would have obtained another proper semifield containing F; and this system has the stronger associativity property that $(ab)c = a(bc)$, whenever <u>any one</u> of a,b,c is in F. Let us call the latter system W.

We will see in Section 6.1 that a proper semifield must contain at least 16 elements.

<u>2.3. Pre-semifields</u>. We say the system S is a pre-semifield if it satisfies all the axioms for a semifield, except possibly A4; i.e., it need not have a multiplicative identity.

A simple example of a pre-semifield can be derived from a field F which has more than one automorphism. In fact, let σ be an automorphism, not the identity, and define $x \circ y = (xy)^\sigma$. Then $(F,+,\circ)$ is a pre-semifield, and it has no identity; for $1 \circ 1 = 1$ implies that 1 must be the identity if any exists, yet $1 \circ y = y^\sigma \neq y$ for some y.

We will see in Section IV that a finite pre-semifield can be thought of as a three-dimensional array of integers, and that a semifield can be constructed from a pre-semifield in several ways.

2.4. The additive group. It is easy to show that the additive group of a pre-semifield S must be commutative. By the distributive laws, $(ac + ad) + (bc + bd) = (a + b)(c + d) = (ac + bc) + (ad + bd)$. Therefore, by A1, $ad + bc = bc + ad$, and any elements which can be written as products commute under addition. But by A2 and finiteness, any element of S can be written as a product; therefore, the additive group is Abelian.

Another simple argument shows that the additive group is *elementary* Abelian. In fact, let $a \neq 0$, and let p be the additive order of a. Then p must be a prime number, since $(na)(ma) = (nm)a$ for integers n,m. The fact that every nonzero element has prime order suffices to show that the group is elementary Abelian, and that all nonzero elements have the same prime order p. This number p is called the *characteristic* of the pre-semifield.

2.5. Vector space representation. Let S be a pre-semifield, and let F be the field $GF(p)$, where p is the characteristic of S. Then we can consider the elements of F to be "scalars," and S is a vector space over F. In particular, S must have p^n elements, where n is the dimension of S over F.

The simple observations just made in the preceding paragraph are surprisingly useful, since many of the concepts of semifields and pre-semifields are fruitfully translated into vector space terminology. This will be exploited further in Section IV. We can, for example, rephrase the distributive laws as follows: Let a be a nonzero element of S; we define the functions L_a and R_a as follows:

$$b\, L_a = ab; \qquad b\, R_a = ba. \tag{2.4}$$

Then the distributive laws A3 are equivalent to the statement that L_a and R_a are linear transformations of the vector space into itself. Furthermore, the axiom A2 states that these transformations are all nonsingular, if $a \neq 0$. Therefore, L_a and R_a can be represented as nonsingular matrices, with elements in $GF(p)$.

<u>2.6. Nuclei</u>. Various special subsystems are defined for a semifield S, indicating degrees of associativity. The most important of these are the following:

The <u>left nucleus</u> N_ℓ: $\{x \mid (xa)b = x(ab),\ a,b \in S\}$.

The <u>middle nucleus</u> N_m: $\{x \mid (ax)b = a(xb),\ a,b \in S\}$.

The <u>right nucleus</u> N_r: $\{x \mid (ab)x = a(bx),\ a,b \in S\}$.

The <u>nucleus</u> N is the intersection of the left, middle, and right nuclei. The field $GF(p)$, where p is the characteristic of S, is obviously always part of the nucleus.

In many cases, the nuclei are all trivial, i.e., equal to $GF(p)$. This is the case for the system V in Section 2.2. But the system W described in that same section has nucleus $F = GF(4)$.

It is easy to verify that each of the nuclei is actually a field. Furthermore, S is a vector space over any of its nuclei: it is a left vector space over N_ℓ, N_m, and N; it is a right vector space over N_m, N_r, and N. The operations L_a and R_a defined by equation (2.4) are not necessarily linear transformations over the nuclei; but R_a is a linear transformation over N_ℓ, and L_a is a linear transformation over N_r, when S is regarded as a left, right vector space, respectively.

III. PROJECTIVE PLANES. ISOTOPY.

Perhaps the major application of semifields today is for the construction of combinatorial designs, and of projective planes in particular. Every proper semifield determines a non-Desarguesian projective plane. In this section we discuss projective planes, and the question whether two semifields coordinatize the same plane.

3.1. Homogeneous coordinates. Let π be a projective plane, and let T be a ternary ring coordinatizing π, as in [1]. We write the ternary operation $a\cdot b\circ c$, and write as usual $a\cdot b\circ 0 = ab$, $a\cdot 1\circ c = a+c$. It is convenient to introduce a new type of coordinate system, for simplicity in notation, as follows:

	Homogeneous Notation	Notation in [1]
Points:	$(0,0,1)$	(∞)
	$(0,1,a)$, $a \in T$	(a)
	$(1,a,b)$, $a,b \in T$	(a,b)
Lines:	$[0,0,1]$	$L(\infty)$
	$[0,1,a]$, $a \in T$	$x + a = 0$
	$[1,a,b]$, $a,b \in T$	$y = x\cdot a\circ b$

The principal feature of this notation is that the point (x_1,x_2,x_3) is incident with the line $[y_1,y_2,y_3]$ if and only if

$$y_1x_3 = x_2\cdot y_2\circ x_1y_3. \qquad (3.1)$$

Since y_1 and x_1 must equal 0 or 1, the meaning of the "multiplications" y_1x_3 and x_1y_3 is clear in this relation. The fact that incidence can be expressed in a single formula means that many special cases can often be eliminated, when carrying out proofs. The symmetry of this definition also makes the duality relation clear.

We let $A{:}B$ denote the line joining points A and B, and let $L \cap M$ denote the point common to lines L and M. Furthermore, we define the elements $a^{\sim}$ and ${}^{\sim}a$ by the equations

$$(a^{\sim}) + a = 0, \qquad a + ({}^{\sim}a) = 0. \tag{3.2}$$

The following formulas can be easily verified:

$$ {}^{\sim}(a^{\sim}) = ({}^{\sim}a)^{\sim} = a \tag{3.3}$$

$$\begin{array}{rlrl}
[0,0,1] & = (0,0,1){:}(0,1,0) & (0,0,1) & = [0,0,1]\cap[0,1,0] \\
[0,1,{}^{\sim}a] & = (0,0,1){:}(1,a,b) & (0,1,a) & = [0,0,1]\cap[1,a,b] \\
[1,a,b] & = (0,1,a){:}(1,0,b) & (1,a,b) & = [0,1,{}^{\sim}a]\cap[1,0,b]
\end{array}$$

3.2. Isotopes. Let T and T_1 be ternary rings. An *isotopism* from T_1 onto T is a set of three functions (F,G,H), each being 1-1 correspondences from T_1 to T, such that

$$\begin{aligned} &(0)H = 0, \\ &(a\cdot b\circ c)H = (aF)\cdot(bG)\circ(cH), \text{ for all } a,b,c \in T . \end{aligned} \tag{3.4}$$

Theorem 3.2.1. *Let* (F,G,H) *be an isotopism from* T_1 *to* T. *Then*

$$H = F\,\psi = G\,\phi \tag{3.5}$$

where $\phi = L_{1F}$, $\psi = R_{1G}$, *and where* L and R *denote left and right multiplication in the ring* T.

Proof: Set $c = 0$ in formula (3.4) to obtain

$$(ab)H = (aF)(bG).$$

In particular,

$$aH = (aF)(1G) = aFR_{1G} = aF\psi,$$

$$bH = (1F)(bG) = bGL_{1F} = bG\phi. \quad \|$$

<u>Theorem 3.2.2</u>. <u>Let</u> T <u>be a ternary ring with</u> n <u>elements. The number of nonisomorphic ternary rings isotopic to</u> T <u>is at most</u> $(n-1)^2$.

Proof: Let T_1 and T_2 be ternary rings isotopic to T, under the functions $(F_1, G_1, H_1)\colon T_1 \to T$ and $(F_2, G_2, H_2)\colon T_2 \to T$. Let the ternary operation of R_1 be denoted by (a,b,c) and let the ternary operation for R_2 be denoted by $[a,b,c]$. We will show that if $1F_1 = 1F_2 = y$, and if $1G_1 = 1G_2 = z$, then T_1 and T_2 are isomorphic. Since there are $(n-1)$ choices for $1F_1$ and $(n-1)$ independent choices for $1G_1$, there will be at most $(n-1)^2$ nonisomorphic ternary rings, as claimed.

By Theorem 3.2.1, we have

$$F_1\psi = G_1\phi = H_1,$$

$$F_2\psi = G_2\phi = H_2,$$

where $\phi = L_y$, $\psi = R_z$. Hence,

$$F_1F_2^{-1} = G_1G_2^{-1} = H_1H_2^{-1} = \alpha.$$

Now α is the required isomorphism, since

$$[a\alpha, b\alpha, c\alpha] = (a\alpha F_2 \cdot b\alpha G_2 \circ c\alpha H_2)H_2^{-1} = (aF_1 \cdot bG_1 \circ cH_1)H_1^{-1}\alpha = (a,b,c)\alpha. \quad \|$$

We note that the limit $(n-1)^2$ is best possible; there is a ternary ring with 32 elements which has 31^2 distinct isotopic ternary rings.

(See Section V.) But if T is a field, we have the other extreme where all isotopic rings are isomorphic to T.

Theorem 3.2.3. Let T *be a ternary ring, and let* y,z *be nonzero elements of* T. *Let* $\phi = L_y$, $\psi = R_z$, *and* $F = \psi^{-1}$, $G = \phi^{-1}$, *and let* T_1 *be the system consisting of the elements of* T *with a new ternary operation defined as follows:*

$$(a,b,c) \;=\; aF \cdot bG \circ c. \tag{3.6}$$

Then T_1 *is also a ternary ring, having the identity element* yz. T_1 *is isotopic to* T, *and furthermore all ternary rings isotopic to* T *can be constructed in this way (up to isomorphism).*

Proof: Note that F and G are well-defined since y and z are nonzero. The latter part of this theorem follows from the preceding theorem; we must show only that (a,b,c) satisfies the requirements for a ternary ring.

I. $(0,b,c) = 0\cdot bG\circ c = c = aF\cdot 0\circ c = (a,0,c)$.

II. $(yz,b,0) = y\cdot bG\circ 0 = y(bG) = bG\phi = b$.

$(a,yz,0) = aF\cdot z\circ 0 = (aF)z = aF\psi = a$.

III. We can solve $(a,b,x) = c$ uniquely for x, since $aF\cdot bG\circ x = c$ is uniquely solvable for x.

IV. We can solve $(x,b_1,c_1) = (x,b_2,c_2)$ uniquely for x, if $b_1 \neq b_2$, since $xF\cdot b_1G\circ c_1 = xF\cdot b_2G\circ c_2$ is uniquely solvable for xF.

V. Finally, we can solve $(a_1,w,x) = c_1$, $(a_2,w,x) = c_2$ uniquely for (w,x), if $a_1 \neq a_2$, since we can solve $a_1F\cdot wG\circ x = c_1$, $a_2F\cdot wG\circ x = c_2$ uniquely for (wG, x). $\blacksquare$

Theorem 3.2.3 is essentially a converse to theorem 3.2.1, for it says that the relations $(0)H = 0$ and $H = F\psi = G\phi$ are sufficient to construct a new ternary operation; no stronger condition can be derived from general isotopy. Theorem 3.2.3 also provides a convenient way to calculate all ternary rings isotopic to a given one. An isotopism where H is the identity is called a "principal isotope."

3.3. The significance of isotopy in geometry. An isomorphism α between projective planes is a 1-1 correspondence between points and lines which preserves incidence; i.e., point P is on line L if and only if point $P\alpha$ is on line $L\alpha$. An automorphism of a plane is commonly called a collineation.

Theorem 3.3.1. Let π and π' be projective planes and let α be an isomorphism from π' onto π such that

$$\begin{aligned}(0,0,1)\alpha &= (0,0,1)\\(0,1,0)\alpha &= (0,1,0)\\(1,0,0)\alpha &= (1,0,0).\end{aligned}$$

Then the ternary rings of π and π' are isotopic.

Proof: $[0,0,1]\alpha = (0,0,1)\alpha:(0,1,0)\alpha = [0,0,1]$. Therefore, $(0,1,a)\alpha$ lies on $[0,0,1]$, and there must be a 1-1 correspondence G such that

$$(0,1,a)\alpha = (0,1,aG). \tag{3.7}$$

Similarly we find $[0,1,0]\alpha = [0,1,0]$, $[1,0,0]\alpha = [1,0,0]$; hence there are 1-1 correspondences H,F such that

$$\begin{aligned}(1,0,b)\alpha &= (1,0,bH).\\(1,a,0)\alpha &= (1,aF,0).\end{aligned} \tag{3.8}$$

We can now calculate the images of all lines:

$$\begin{aligned} [0,1,{\sim}a]\alpha &= (0,0,1)\alpha:(1,a,0)\alpha = [0,1,{\sim}(aF)] \\ [1,a,b]\alpha &= (0,1,a)\alpha:(1,0,b)\alpha = [1,aG,bH]. \end{aligned} \tag{3.9}$$

Finally, we find the image of every point:

$$(1,a,b)\alpha = [0,1,{\sim}a]\alpha \cap [1,0,b]\alpha = (1,aF,bH). \tag{3.10}$$

Now we can derive the desired law:

$$(1,x_2,x_3) \in [1,y_2,y_3] \;\leftrightarrow\; (1,x_2,x_3)\alpha \in [1,y_2,y_3]\alpha \tag{3.11}$$

i.e. $$x_3 = x_2 \cdot y_2 \circ y_3 \;\leftrightarrow\; x_3H = x_2F \cdot y_2G \circ y_3H,$$

and this is precisely the relation used for isotopy. ▌

Theorem 3.3.2. (Converse of Theorem 3.3.1). Let (F,G,H) *be an isotopy from a ternary ring* T' *to* T, *and let* π',π *be the corresponding planes. Define* α *by equations (3.7)-(3.10); then* α *is an isomorphism between* π' *and* π.

Proof: Since α is 1-1 and onto, and since (3.11) holds directly from the law of isotopy, there are only a few cases to consider.

$$\begin{array}{rlcl} \text{I.} & (1,x_2,x_3) \in [0,y_2,y_3] & \leftrightarrow & y_2 = 1 \text{ and } y_3^{\sim} = x_2 \\ & (1,x_2,x_3)\alpha \in [0,y_2,y_3]\alpha & \leftrightarrow & y_2 = 1 \text{ and } y_3^{\sim}F = x_2F. \\ \text{II.} & (0,x_2,x_3) \in [1,y_2,y_3] & \leftrightarrow & x_2 = 1 \text{ and } x_3 = y_2 \\ & (0,x_2,x_3)\alpha \in [1,y_2,y_3]\alpha & \leftrightarrow & x_2 = 1 \text{ and } x_3G = y_2G. \\ \text{III.} & (0,x_2,x_3) \in [0,y_2,y_3] & \leftrightarrow & x_2 = 0 \text{ or } y_2 = 0 \\ & (0,x_2,x_3)\alpha \in [0,y_2,y_3]\alpha & \leftrightarrow & x_2 = 0 \text{ or } y_2 = 0. \end{array}$$ ▌

We define autotopism in an obvious way, as an isotopism of a ternary ring onto itself. If (F,G,H) and (F',G',H') are autotopisms,

$$(a\cdot b\circ c)HH' = (aF\cdot bG\circ cH)H' = aFF'\cdot bGG'\circ cHH',$$

so we define the product of two autotopisms as

$$(F,G,H)(F',G',H') = (FF',GG',HH'). \tag{3.12}$$

An automorphism is the special case (F,F,F) of an autotopism where all three permutations are equal.

Corollary 3.3.3. All isotopic ternary rings coordinatize the same projective plane. The collineations of a projective plane which fix $(0,0,1)$, $(0,1,0)$, *and* $(1,0,0)$ *form a group isomorphic to the group of autotopisms of the ternary ring.* ▮

Theorem 3.3.4. Let T *be a ternary ring with* n *elements, and let* h *be the order of the autotopism group of* T. *Then*

$$(n-1)^2 = \sum_{T'} \frac{h}{k(T')} \tag{3.13}$$

where T' *ranges over all nonisomorphic ternary rings isotopic to* T, *and where* $k(T')$ *is the number of automorphisms of* T'.

Proof: Let y,z range over the nonzero elements of T, and consider the $(n-1)^2$ ternary rings constructed in Theorem 3.2.3. If T' is any of these ternary rings, we will show that there are exactly $h/k(T')$ ternary rings of the set which are isomorphic to T', and this will prove formula (3.13).

We need only show that $h/k(T)$ of the ternary rings are isomorphic to T, because the autotopism group of T' is conjugate to the autotopism group of T and therefore has the same order, and because the $(n-1)^2$ ternary rings formed from T' are isomorphic in some order to the $(n-1)^2$ ternary rings formed from T (using Theorem 3.2.2, since the rings are determined by $1F$ and $1G$).

Let α be an isomorphism from T to the ring $T'(R_z^{-1}, L_y^{-1}, 1)$. There are $k(T)$ such isomorphisms. Then

$$(a \cdot b \circ c)\alpha = (a\alpha, b\alpha, c\alpha) = a\alpha R_z^{-1} \cdot b\alpha L_y^{-1} \circ c\alpha,$$

i.e., $(\alpha R_z^{-1}, \alpha L_y^{-1}, \alpha)$ is an autotopism. By Theorem 3.2.1, every autotopism is of this form, and defines y and z; therefore, if r of the pairs y,z yield isomorphic rings, there are $h = rk(T)$ autotopisms. ▌

<u>3.4. Isotopy of semifields.</u> A semifield is a particular type of ternary ring, where $a \cdot b \circ c = ab + c$ and axioms A1-A4 hold. We now specialize the material of the preceding sections to the case of semifields.

<u>Theorem 3.4.1.</u> *Let S be a semifield of characteristic p. All ternary rings isotopic to S are semifields. (F,G,H) is an isotopism from S' to S if and only if F, G, and H are nonsingular linear transformations from S' to S over $GF(p)$, satisfying the condition*

$$(ab)H = (aF)(bG). \tag{3.14}$$

Proof: Suppose S' is isotopic to S, and $(a \cdot b \circ c)H = (aF)(bG) + cH$. Then $(a \cdot 1 \circ c)H = (aF)(1G) + cH = aH + cH$, so H is an isomorphism between the addition of S' and the addition of S, i.e., H is a nonsingular linear transformation over $GF(p)$. Now $H = F\psi = G\phi$

by Theorem 3.2.1, where ψ and ϕ, being functions of left and right multiplication, are nonsingular linear transformations over $GF(p)$ of S into itself; hence F and G are also nonsingular linear transformations. Properties A1 through A4 are now verified immediately, as is the converse portion of the theorem. ‖

Theorem 3.4.2. Every collineation of a plane coordinatized by a proper semifield fixes $(0,0,1)$ and $[0,0,1]$.

Proof: This theorem is well known, but its proof requires the development of more geometrical tools then are appropriate here. Proofs may be found in [2] and [3]. ‖

The previous theorem holds primarily because any semifield already has a great number of collineations, and if it were permitted to move the point $(0,0,1)$ or $[0,0,1]$ it would have so many more collineations, it would become Desarguesian.

The standard collineations, holding in any semifield plane, are the translations $\tau(h,k)$ and the generalized shears $\sigma(h,k)$, defined for all h,k in the semifield as follows:

$$\begin{aligned}(x_1,x_2,x_3)\tau(h,k) &= (x_1,x_2+x_1h,x_3+x_1k)\\ [y_1,y_2,y_3]\tau(h,k) &= [y_1,y_2,y_3-hy_2+y_1k]\\ (x_1,x_2,x_3)\sigma(h,k) &= (x_1,x_2,x_3+x_2h+x_1k)\\ [y_1,y_2,y_3]\sigma(h,k) &= [y_1,y_2+y_1h,y_3+y_1k]\end{aligned} \tag{3.15}$$

The proof that these are collineations is very simple with homogeneous coordinates: $y_1x_3 = x_2y_2 + x_1y_3$

if and only if $y_1(x_3+x_1k) = (x_2+x_1h)y_2 + x_1(y_3-hy_2+y_1k)$

if and only if $y_1(x_3+x_2h+x_1k) = x_2(y_2+y_1h) + x_1(y_3+y_1k)$,

remembering that x_1 and y_1 are restricted to be 0 or 1.

The following relations are easily computed, using the formulas already derived; here $\alpha(F,G,H)$ represents a collineation corresponding to an autotopism.

$$\begin{aligned}\tau(0,k) &= \sigma(0,k)\\ \tau(h,k)\tau(h',k') &= \tau(h+h',k+k')\\ \sigma(h,k)\sigma(h',k') &= \sigma(h+h',k+k')\\ \alpha(F,G,H)\alpha(F',G',H') &= \alpha(FF',GG',HH')\\ \tau(h,k)^{-1}\sigma(\ell,m)\tau(h,k) &= \sigma(\ell,m-h\ell)\\ \sigma(h,k)^{-1}\tau(\ell,m)\sigma(h,k) &= \tau(\ell,m+\ell h)\\ \alpha(F,G,H)^{-1}\tau(h,k)\alpha(F,G,H) &= \tau(hF,kH)\\ \alpha(F,G,H)^{-1}\sigma(h,k)\alpha(F,G,H) &= \sigma(hG,kH)\end{aligned} \tag{3.16}$$

Theorem 3.4.3. Two semifields coordinatize the same plane if and only if they are isotopic.

Proof: If β is an isomorphism between two semifield planes, then $(0,0,1)\beta = (0,0,1)$ and $[0,0,1]\beta = [0,0,1]$ because this point and line are characterized by Theorem 3.4.2. Hence $(0,1,0)\beta = (0,1,a)$, and $(1,0,0)\beta = (1,b,c)$. Let $\alpha = \beta\sigma(-a,0)\tau(-b,ba-c)$. Then

$$(0,0,1)\alpha = (0,0,1); \quad (1,0,0)\alpha = (1,0,0); \quad (0,1,0)\alpha = (0,1,0);$$

and Theorem 3.3.1 applies. The converse is part of Corollary 3.3.3. ▮

Theorem 3.4.4. Let G *be the collineation group of a semifield plane; let* T *be the subgroup of translations,* S *the subgroup of generalized shears,* H *the subgroup corresponding to autotopisms, and* A *the elementary Abelian additive group of the semifield. Then we have the following normal series for* G:

$$I \lhd T \cap S \lhd T \lhd T \cup S \lhd T \cup S \cup H = G.$$

The quotient groups are equal respectively to A, A, A, *and* H.

Proof: This follows from formulas (3.16), and the fact that $T \cup S \cup H = G$ as proved in the preceding theorem. ‖

Theorems 3.4.3 and 3.4.4 are well known; they indicate how important isotopy is when considering semifield planes. The use of homogeneous coordinates simplifies and clarifies previous proofs of these theorems.

3.5. Nonlinear isotopes. One might pose an interesting problem here, concerning whether linearity of F,G,H is really a necessary condition for constructing semifields. Suppose we have a semifield S and we have defined a new multiplication * on S by the formula

$$(a*b)H = (aF)(bG).$$

We require that F,G,H be permutations of S, and that

$$\begin{aligned} (a+b)*c &= a*c + b*c, \\ c*(a+b) &= c*a + c*b, \qquad (3.17) \\ 1*a &= a*1 = a. \end{aligned}$$

Question: Does this imply that F,G,H are linear?

The following theorem does not settle this question by any means, but it does provide some insight into the matter. We say an element $x \in S$ is <u>noncentral</u> if $xy = yx$ implies that

$$y = r + sx, \quad r,s \in GF(p).$$

<u>Theorem 3.5.1. If</u> S <u>has characteristic 2, and if every element of</u> $S - GF(2)$ <u>is noncentral, then the functions</u> F,G,H <u>of</u> (3.17) <u>must be linear.</u>

Proof: The conditions imply that $aH^{-1} = aF^{-1} * 1G^{-1} = 1F^{-1} * aG^{-1}$, so we can write $H = PF = QG$ where P and Q are linear. Thus

$$(aH)(bH) = (aP * bQ)H. \qquad (3.18)$$

Let x be such that $xH = 1$. Then since $(aH)(xH) = (xH)(aH)$, equation (3.18) shows us that $aP*xQ = xP*aQ$, for all $a \in S$. Adding $aP*aQ$ to each side, we obtain $aP * (a+x)Q = (a+x)P * aQ$; hence

$$(aH)((a+x)H) = ((a+x)H)(aH),$$

i.e., aH commutes with $(a+x)H$. This is the relation which we will use in order to show that H is linear.

We can at least prove that $(0)H = 0$, since $(0H)(aH) = (0P*aQ)H = (0)H$ for all a. Therefore $x \neq 0$. We will now show that

$$(a + x)H = aH + xH = aH + 1, \text{ for all } a. \qquad (3.19)$$

If $a = 0$ or if $a = x$, this is obviously true (using the fact that $x + x = 0$). Otherwise aH is noncentral, by hypothesis. By the definition, this implies that $(a+x)H = 0$ or 1 or aH or $aH + 1$.

The first three possibilities are clearly impossible, so (3.19) is established.

Since $H = G\phi$, where ϕ is linear, we have

$$(a+x)G = (a+x)H\phi^{-1} = aH\phi^{-1} + xH\phi^{-1} = aG + xG.$$

Finally, then, let a,b be arbitrary nonzero elements of S; define c,d such that $a = c*x$, $b = c*d$. Then

$$\begin{aligned}(a+b)H = (c*(x+d))H &= (cF)((x+d)G) \\ &= (cF)(xG + dG) \\ &= (cF)(xG) + (cF)(dG) = aH + bH.\end{aligned}$$

Thus H is linear and therefore so are F and G. ▯

Remark. The above theorem is not trivial, for there exist systems in which the hypothesis is satisfied (e.g., V in Section 2.2). Furthermore, some hypothesis is necessary for the theorem, because there are examples in which F,G,H are <u>not</u> linear. Such an example is the field $S = GF(8)$. Let x be a primitive element with minimum polynomial x^3+x^2+1; then define H by

$$(\ x^i \pmod{x^3+x^2+1}\ )H = x^i \pmod{x^3+x+1}.$$

If we write $(a*b)H = (aH)(bH)$, the equations (3.17) are satisfied, but $(1 + x^2)H = 1 + x \neq 1 + x^2 = 1H + x^2H$.

Such nonlinear isotopies do not have geometric significance, however, according to Theorem 3.4.1, and they will not be considered further in this paper.

IV. NONSINGULAR HYPERCUBES.

In this section we turn to another way of looking at semifields and their isotopisms, where we think of 3-dimensional matrices. The 3-dimensional representation allows us to see several symmetries in the situation which are not otherwise apparent; and it also allows us to work with pre-semifields, when it is convenient.

4.1. Hypercubes and their elementary operations.

An m-dimensional hypercube A, $m \geqq 1$, is an array of n^m elements belonging to a field; the elements are denoted by $A_{ij\ldots r}$ where there are m subscripts, and each subscript varies from 1 to n. We consider n to be a fixed integer throughout the entire discussion.

It is easy to devise extremely cumbersome notations for such systems, so an attempt will be made to keep the notation as simple as possible.

Let σ be a permutation on the elements $1,2,\ldots m$. Then A^σ will represent the m-cube A with subscripts permuted by σ; i.e., the $k\sigma$-th subscript of A^σ is the k-th subscript of A. This is a generalization of the concept of transposition of matrices; if A is a matrix $(m = 2)$, $A^T = A^{(12)}$. In the 3-dimensional case, if $A = (A_{ijk})$, then if $B = A^{(123)} = (B_{ijk})$ we have $B_{ijk} = A_{jki}$ for all i,j,k. We also have the general law $(A^\sigma)^\tau = A^{\sigma\tau}$.

If $m > 1$, we form $(m-1)$-dimensional subcubes of an m-cube A, denoted by $A^{[t]}$, as follows:

$$B = A^{[t]} = (B_{ij\ldots r}) \leftrightarrow B_{ij\ldots r} = A_{tij\ldots r}. \tag{4.1}$$

Here $1 \leqq t \leqq n$. Additional subcubes which hold other positions fixed can be formed by combining the operations A^σ and $A^{[t]}$ described here; actually all conceivable subcubes can be obtained in this manner.

4.2. Sums and products of hypercubes. The sum of two m-cubes A and B is simply defined as the m-cube consisting of the sums of the components:

$$C = A + B = (C_{ij\ldots r}) \leftrightarrow A_{ij\ldots r} + B_{ij\ldots r} = C_{ij\ldots r} \tag{4.2}$$

The products of a p-cube B times an m-cube A, yielding an $(m+p-2)$-cube, can be defined in the same way as the tensor dot product of tensors, as follows:

$$C = B \overset{1}{\times} A = (C_{i\ldots jk\ldots r}) \leftrightarrow C_{i\ldots jk\ldots r} = \sum_t B_{i\ldots jt}A_{tk\ldots r}.$$

Actually, n of such products can be defined, and we write in general

$$C = B \overset{\ell}{\times} A = (C_{i\ldots jk\ldots qr\ldots s}) \leftrightarrow$$

$$C_{i\ldots jk\ldots qr\ldots s} = \sum_t B_{k\ldots qt}A_{i\ldots jtr\ldots s} \tag{4.3}$$

where "i...j" represents $\ell-1$ subscripts, "r...s" represents $n-\ell$.

Notice that $A^{[t]} = B \overset{1}{\times} A$, where B is the 1-cube $(B_i) = (\delta_{it})$. We also see that if A,B,C are matrices, $B \overset{1}{\times} A = BA$, $C \overset{2}{\times} A = AC^T$, expressing the familiar fact that premultiplication corresponds to operating on rows of a matrix, postmultiplication corresponds to operating on columns.

The associative law for matrix multiplication, $(BA)C^T = B(AC^T)$, can be written in the form $C \overset{2}{\times} (B \overset{1}{\times} A) = B \overset{1}{\times} (C \overset{2}{\times} A)$. This relation is a special case of a general rule for products of hypercubes.

Theorem 4.2.1. (Generalized associative law). If $u < v$, and if the dimension of C is $f + 2$, then

$$C \overset{u}{\times} (B \overset{v}{\times} A) = B \overset{v+f}{\times} (C \overset{u}{\times} A) \tag{4.4}$$

Proof: Let the dimensions of A and B be m and p, respectively. The general element $D_{i\ldots jk\ldots qr\ldots sw\ldots xy\ldots z}$ of the product $C \overset{u}{\times} (B \overset{v}{\times} A)$, where $i\ldots j$ represents $u-1$ subscripts, $k\ldots q$ represents $f+1$ subscripts, $r\ldots s$ respresents $v-u-1$ subscripts, $w\ldots x$ represents $p-1$ subscripts, and $y\ldots z$ represents $m-v$ subscripts, is, by definition,

$$\sum_t C_{k\ldots qt}\Big(\sum_h B_{w\ldots xh}A_{i\ldots jtr\ldots shy\ldots z}\Big).$$

The same general element of the right-hand product is

$$\sum_t B_{w\ldots xt}\Big(\sum_h C_{k\ldots qh}A_{i\ldots jhr\ldots sty\ldots z}\Big),$$

and these are clearly equal. ▮

Corollary 4.2.2. *Let* m *be the dimension of* A, *and let* $B_1, B_2, \dots B_m$ *be matrices. Then the* m *multiplications* $B_i \overset{i}{\times} A$ *can be performed on* A *in any order, i.e., if* β *is a permutation,*

$$\begin{aligned} &B_1 \overset{1}{\times} (B_2 \overset{2}{\times} (\cdots (B_m \overset{m}{\times} A)\cdots)) \\ = \quad &B_{1\beta} \overset{1\beta}{\times} (B_{2\beta} \overset{2\beta}{\times} (\cdots (B_{m\beta} \overset{m\beta}{\times} A)\cdots)). \end{aligned} \tag{4.5}$$

Proof: If the dimension is 2, then f in Theorem 4.2.1 is always zero, so this result follows by repeated application of that theorem. ‖

We will write

$$[B_1, B_2, \dots B_m] \times A$$

to denote the multiplication operations expressed in equation (4.5).

The following law generalizes the matrix equation $(CA)^T = A^T C^T$:

Theorem 4.2.3. *If* C *is a matrix, and if* σ *is a permutation,*

$$(C \overset{u}{\times} A)^\sigma \quad = \quad C \overset{u\sigma}{\times} A^\sigma \tag{4.6}$$

Proof: Let the subscripts be $i_1, i_2, \dots i_m$. The desired result follows from the formulas

$$\begin{aligned} (A^\sigma)_{i_1 i_2 \dots i_m} &= A_{i_{1\sigma} i_{2\sigma} \dots i_{m\sigma}} \\ (C \overset{u}{\times} A)_{i_1 i_2 \dots i_m} &= \sum_t C_{i_u t} A_{i_1 \dots i_{u-1} t i_{u+1} \dots i_m} \\ (C \overset{u\sigma}{\times} A^\sigma)_{i_1 i_2 \dots i_m} &= \sum_t C_{i_{u\sigma} t} A_{i_{1\sigma} \dots i_{(u-1)\sigma} t i_{(u+1)\sigma} \dots i_{m\sigma}} . \end{aligned} \quad ‖$$

Several other associative laws can be formulated to include the cases not meeting the hypothesis of Theorem 4.2.1; i.e., when we are given a product $C \overset{u}{\times} (B \overset{v}{\times} A)$ such that $u \geqq v$ and that $u+2-\dim C \leqq v$. But only one of these is of concern to us here:

Theorem 4.2.4.

$$[C_1,C_2,\ldots,C_m] \times ([B_1,B_2,\ldots,B_m] \times A) = [C_1B_1,C_2B_2,\ldots C_mB_m] \times A. \qquad (4.7)$$

Proof: In lieu of Theorems 4.2.1 and 4.2.3, it suffices to show that $(C \overset{1}{\times} (B \overset{1}{\times} A) = (CB) \overset{1}{\times} A$ when C,B are matrices. The general element $D_{ij\ldots r}$ of the left hand side is

$$\sum_t C_{it}(\sum_h B_{th}A_{hj\ldots r});$$

on the right hand side it is

$$\sum_t (\sum_h C_{ih}B_{ht})A_{tj\ldots r},$$

and these are equal. ‖

4.3. Nonsingular hypercubes. The concept of a nonsingular matrix can be generalized to m-cubes in a significant way. The definition proceeds inductively: We say that a vector, or 1-cube, A is singular if and only if $A = 0$. For $m > 1$, we say that an m-cube A is singular if and only if there exists a nonsingular vector B such that $B \overset{1}{\times} A$ is singular.

Another way to state this definition is that an m-cube A is <u>nonsingular</u> if the following condition is satisfied:

$$\text{"}x_1A^{[1]}+x_2A^{[2]}+ \dots + x_nA^{[n]} \text{ is singular implies that}$$

$$x_1 = x_2 = \dots = x_n = 0.\text{"}$$

In other words, any nonzero linear combination of the subcubes $A^{[t]}$ must be nonsingular. This definition certainly includes the ordinary definition of a nonsingular matrix, for in the special case $m = 2$ it says that the rows of A are linearly independent.

<u>Theorem 4.3.1</u>. <u>Let</u> A <u>be an m-cube, and let</u> σ <u>be a permutation of</u> $\{1,2,\dots,m\}$. <u>Then</u> A <u>is nonsingular if and only if</u> A^σ <u>is nonsingular</u>.

Proof: We use induction on m. For $m = 1$ it is trivial, and for $m = 2$ it is a special case of the theorem that "row rank equals column rank of a matrix." We assume, therefore, that $m > 2$.

Without loss of generality we may assume that σ is of the form (12) or $(23\cdots m)$, since these two permutations generate all of the $m!$ permutations.

If $\sigma = (12)$, we argue as follows, where B and C denote vectors:

$$A \text{ is nonsingular} \leftrightarrow (C \neq 0 \rightarrow C \overset{1}{\times} A \text{ is nonsingular})$$

$$\leftrightarrow (C \neq 0 \rightarrow (B \neq 0 \rightarrow B \overset{1}{\times}(C \overset{1}{\times} A) \text{ is nonsingular})) \text{ since } m > 2$$

$$\leftrightarrow (B \neq 0 \rightarrow (C \neq 0 \rightarrow B \overset{1}{\times}(C \overset{1}{\times} A) \text{ is nonsingular}))$$

$$\leftrightarrow (B \neq 0 \rightarrow (C \neq 0 \rightarrow C \overset{1}{\times}(B \overset{2}{\times} A) \text{ is nonsingular})) \text{ by Th. 4.2.1}$$

$$\leftrightarrow (B \neq 0 \rightarrow B \overset{2}{\times} A \text{ is nonsingular})$$

$$\leftrightarrow (B \neq 0 \rightarrow B \overset{1}{\times} A^\sigma \text{ is nonsingular}) \text{ since } B \overset{1}{\times} A^{(12)} = B \overset{2}{\times} A$$

$$\leftrightarrow A^\sigma \text{ is nonsingular.}$$

On the other hand if $\sigma = (23\cdots m)$, let $\tau = (12\cdots(m-1))$. Then since $(B \overset{1}{\times} A)^{\tau} = B \overset{1}{\times} (A^{\sigma})$ we argue as follows:

$$\begin{aligned} A \text{ is nonsingular} &\leftrightarrow (B \neq 0 \to B \overset{1}{\times} A \text{ is nonsingular}) \\ &\leftrightarrow (B \neq 0 \to (B \overset{1}{\times} A)^{\tau} \text{ is nonsingular}) \text{ by induction} \\ &\leftrightarrow (B \neq 0 \to B \overset{1}{\times} (A^{\sigma}) \text{ is nonsingular}) \\ &\leftrightarrow A^{\sigma} \text{ is nonsingular.} \quad \| \end{aligned}$$

Theorem 4.3.2. *Let* A *be an m-cube and let* $C_1, C_2, \ldots, C_m$ *be nonsingular matrices. Then* A *is nonsingular if and only if*

$$[C_1, C_2, \ldots, C_m] \times A$$

is nonsingular.

Proof: For $m = 1$ this is merely the definition of a singular matrix. For $m > 1$, it suffices to show that A is singular if, and only if, $C \overset{2}{\times} A$ is singular because of Theorem 4.2.3 and Theorem 4.3.1.

$$\begin{aligned} C \overset{2}{\times} A \text{ is nonsingular} &\leftrightarrow (B \neq 0 \to B \overset{1}{\times} (C \overset{2}{\times} A) \text{ is nonsingular}) \\ &\leftrightarrow (B \neq 0 \to C \overset{1}{\times} (B \overset{1}{\times} A) \text{ is nonsingular}) \text{ by Th. } 4.2.1 \\ &\leftrightarrow (B \neq 0 \to B \overset{1}{\times} A \text{ is nonsingular}) \text{ by induction} \\ &\leftrightarrow A \text{ is nonsingular.} \quad \| \end{aligned}$$

Finally, we define an equivalence relation between m-cubes as follows:

$$A \equiv B \text{ if and only if } A = [C_1, C_2, \ldots, C_m] \times B \tag{4.8}$$

for nonsingular matrices $C_1, C_2, \ldots, C_m$. By Theorem 4.2.4 this is an equivalence relation, and by the preceding theorem this preserves singularity.

4.4. Pre-semifields represented as 3-cubes. We now specialize to the case $m = 3$. Let S be a pre-semifield of characteristic p. According to Section 2.5, S is a vector space over $F = GF(p)$. Let $\{x_1, x_2, \ldots, x_n\}$ be a basis of S over F. We can write the multiplication in terms of the basis elements:

$$x_i \, x_j = \sum_k A_{ijk} \, x_k. \tag{4.9}$$

This gives us a 3-cube, A, with entries in F. A is called a cube corresponding to S.

The multiplication in S is completely determined by the products of the basis elements, according to the distributive laws since

$$(\sum_i b_i x_i)(\sum_j c_j x_j) = \sum_i \sum_j b_i c_j x_i x_j = \sum_i \sum_j \sum_k b_i c_j A_{ijk} x_k. \tag{4.10}$$

Theorem 4.4.1. A cube corresponding to a pre-semifield is nonsingular. Conversely, if A is a nonsingular 3-cube, we can define a pre-semifield S by equation (4.10), with the x_i being formal basis elements.

Proof: Because of our other observations, we need show only that multiplication defined by equation (4.10) satisfies axiom A2 if and only if A is nonsingular. In the proof of Theorem 4.3.1, we have shown that A is nonsingular if and only if

$B \neq 0$ and $C \neq 0 \rightarrow C \overset{1}{\times} (B \overset{1}{\times} A) \neq 0$ (in the case $m = 3$).

But $C \overset{1}{\times} (B \overset{1}{\times} A)$ is the vector $D = (D_k)$ where

$$D_k = \sum_r B_r \sum_s C_s A_{rsk};$$

therefore, this is precisely the condition that

$$a \neq 0 \text{ and } b \neq 0 \rightarrow ab \neq 0. \quad \|$$

We observe that a nonsingular 4-cube would correspond to an algebraic system with a ternary multiplication abc satisfying three distributive laws, and with no "zero-divisors." In general, a nonsingular m-cube will lead to an (m-1)-ary operation.

If we change to a different basis $\{y_1, \ldots, y_n\}$ of S over F, we have

$$y_i = \sum_j C_{ij} x_j, \quad \text{where} \quad C = (C_{ij}) \text{ is a nonsingular matrix.}$$

This introduces a corresponding change in A; let the new cube be B. Then

$$\begin{aligned} y_i y_j = \Big(\sum_r C_{ir} x_r\Big)\Big(\sum_s C_{js} x_s\Big) &= \sum_r \sum_s \sum_k C_{ir} C_{js} A_{rsk} x_k \\ &= \sum_r \sum_s \sum_k \sum_t C_{ir} C_{js} A_{rsk} C^{-1}_{kt} y_k . \end{aligned}$$

Now B_{ijk} is the coefficient of y_k, hence

$$B = [\ C\ ,\ C\ ,\ C^{-T}\ ] \times A, \tag{4.11}$$

where $-T$ denotes inverse transpose. Thus $B \equiv A$.

Let us now consider what would happen if we were to define a new multiplication on the elements of S:

$$(a * b)H = (aF)(bG) \tag{4.12}$$

where F, G, H are arbitrary nonsingular linear transformations of S into itself. This gives us another pre-semifield S', which is said to be isotopic to S.

Let B be a cube corresponding to the derived pre-semifield S'. We may assume S' has the same basis $\{x_1, x_2, \ldots x_n\}$ as S, and that A is the cube for S with this basis. Consider F,G,H as matrices;

i.e.,
$$x_i F = \sum_r F_{ir} x_r . \tag{4.13}$$

Then
$$\begin{aligned} x_i * x_j &= ((x_i F)(x_j G))H^{-1} = ((\sum_r F_{ir} x_r)(\sum_s G_{js} x_s))H^{-1} \\ &= (\sum_r \sum_s \sum_t F_{ir} G_{js} A_{rst} x_t)H^{-1} \\ &= \sum_r \sum_s \sum_t \sum_k F_{ir} G_{js} H^{-1}_{tk} A_{rst} x_k . \end{aligned} \tag{4.14}$$

Therefore
$$B = [\ F\ ,\ G\ ,\ H^{-T}] \times A. \tag{4.15}$$

This proves the following fundamental theorem.

Theorem 4.4.2. Let S and S' be pre-semifields, and let A,A' be any cubes corresponding to S and S'. Then S is isotopic to S' if and only if $A \equiv A'$ (where equivalence is defined in equation (4.8)).

4.5. Semifields and equivalent pre-semifields. The preceding discussion applies to semifields as special cases of pre-semifields. If S is a semifield, we can assume the basis is of the form $\{1,x_2,\ldots,x_n\}$. The corresponding cube A then has a special property.

We write

$$\begin{aligned} A^{r**} &= A[r] \\ A^{*r*} &= (A^{(132)})[r] \\ A^{**r} &= (A^{(123)})[r]. \end{aligned} \qquad (4.16)$$

In other words, A^{r**}, A^{*r*}, A^{**r} are the matrices $(A_{rij}),(A_{jri}),(A_{ijr})$ respectively. A^{r**} is the matrix L_{x_r} of left multiplication by x_r. A^{*r*} is the transpose of the matrix for right multiplication, R_{x_r}. The following formulas are readily verified:

$$\begin{aligned} ([F,G,H] \times A)^{r**} &= F_{r1}(GA^{1**}H^T) + \ldots + F_{rn}(GA^{n**}H^T). \\ ([F,G,H] \times A)^{*r*} &= G_{r1}(HA^{*1*}F^T) + \ldots + G_{rn}(HA^{*n*}F^T). \\ ([F,G,H] \times A)^{**r} &= H_{r1}(FA^{**1}G^T) + \ldots + H_{rn}(FA^{**n}G^T). \end{aligned} \qquad (4.17)$$

We say A is in standard form if $A^{1**} = A^{*1*} = I$.

Theorem 4.5.1. Let S *be a semifield with basis* $\{1,x_2,\ldots,x_n\}$; *then the cube corresponding to* S *is nonsingular and in standard form; conversely, every nonsingular 3-cube in standard form yields a semifield, if multiplication is defined by (4.10).*

Proof: Because of Theorem 4.4.1, we must merely observe that standard form is equivalent to axiom A4. But this is obvious, since standard form is nothing but the statement that $L_1 = R_1 = I$. ▮

Theorem 4.5.2. *Let* S, S' *be semifields with corresponding cubes* A, A'. *Then* S *and* S' *coordinatize the same projective plane if and only if* $A \equiv A'$. *Semifield planes are in 1-1 correspondence with equivalence classes of nonsingular 3-cubes*.

Proof: Apply Theorems 4.4.2 and 3.4.3. ▮

Theorem 4.5.3. *If* A *is a cube corresponding to a proper semifield* S, *the autotopism group of* S *is isomorphic to the group of all triples of matrices* (F,G,H) *such that*

$$[\,F\,,\,G\,,\,H\,] \times A = A. \tag{4.18}$$

Proof: This follows from equation (4.15). ▮

We note that the same result holds for any 3-cube B, if $B \equiv A$, even if B is not in standard form.

Now we turn to the question of constructing a semifield from a pre-semifield. This is merely a question of finding three nonsingular matrices such that $[F,G,H] \times A$ is in standard form. Such a construction can be done in several ways; for example:

1. Set $G = (A^{1**})^{-1}$, set $B = G \overset{2}{\times} A$, set $F = (B^{*1*})^{-T}$, $H = I$.

2. Set $F = (A^{*1*})^{-T}$, set $B = F \overset{1}{\times} A$, set $G = (B^{1**})^{-1}$, $H = I$.

3. Set $H = (A^{1**})^{-T}$, $F = A^{1**}(A^{*1*})^{-T}$, $G = I$.

4. Set $H = (A^{*1*})^{-1}$, $G = A^{*1*T}(A^{1**})^{-1}$, $F = I$.

The proofs that these do the job are similar, using equations (4.17). We consider for example method 3:

$$F = \begin{pmatrix} a_{111} & a_{112} & \cdots & a_{11n} \\ a_{121} & a_{122} & & a_{12n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \end{pmatrix} \begin{pmatrix} a_{111} & a_{112} & \cdots & a_{11n} \\ a_{211} & a_{212} & & a_{21n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ * & * & & * \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \end{pmatrix}.$$

Therefore $([F,G,H] \times A)^{1**} = GA^{1**}H^T = I$;

and $([F,G,H] \times A)^{*1*} = HA^{*1*}F^T = I$.

These methods can all be translated into algebraic terms; e.g., here is method 3 in this form:

Theorem 4.5.4. Let $(S,+,\circ)$ be a pre-semifield, and let $u \in S$. Then if we define a new multiplication $*$ by the rule

$$(a \circ u) * (u \circ b) = a \circ b \qquad (4.19)$$

we obtain a semifield $(S,+,*)$ with unit $u \circ u$. ∎

Notice that if $(S,+,\circ)$ is commutative, so is $(S,+,*)$.

Formula (4.19) is really an instance of equation (3.6) which is actually more general. A less symmetric way to write this formula can be read directly from method 3, namely

$$u \circ (a * b) = (u \circ a)R_u^{-1} \circ b,$$

which yields an isomorphic system with unit element u. Method 4 gives another isomorphic system, dual to this one.

Thus, we can obtain semifields from pre-semifields in several ways; in each method we were able to leave F, G, or H equal to the identity. No matter which way we choose, the result is equivalent in the sense that the same projective plane will result.

The extra degree of freedom we seem to have in this standardization process indicates that we should seek a "more standard" standard form, so that it might actually be a canonical form. One idea suggests itself immediately: we could require that $A^{**1} = I$ also.

If S is a field with p^2 elements, it can always be put into the form

$$A^{1**} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad A^{2**} = \begin{pmatrix} 0 & 1 \\ 1 & a \end{pmatrix},$$

if a is chosen such that $x^2 + ax - 1$ is irreducible (mod p).

The field GF(8) can be put into the forms

$$\begin{array}{ccccccc} 100 & 010 & 001 & & 100 & 010 & 001 \\ 010 & 101 & 001 & \text{or} & 010 & 111 & 011 \\ 001 & 011 & 111 & & 001 & 011 & 110 \end{array}$$

where we adopt the convention of writing $A^{1**}, A^{2**}, \ldots$ in this order. It is unclear whether this form is possible in general. These examples show that it is not a canonical form; but if it is true that all nonsingular cubes can be standardized in this way, it would be useful for eliminating cases when constructing the systems.

A way to utilize the extra freedom which has been very fruitful is to adjust $A^{2}**$. We can essentially perform any desired similarity transformation on $A^{2}**$, while the standardization is taking place; therefore, in particular, if the characteristic equation of $A^{2}**$ after standardization is irreducible, we can transform it into a companion matrix and restandardize; this means we are taking the basis of the vector space to be of the form 1, x, x^2, $x(x^2)$, $x(x(x^2))$, etc. Operations of this kind can be exploited when all possible semifields of a given order are to be constructed [see reference 2].

V. TRANSPOSE OF A PLANE.

This section discusses an interesting relationship between some planes coordinatized by semifields. The relationship is somewhat peculiar since it has algebraic significance but does not seem to have any readily seen geometric significance.

5.1. Transpose defined. Let π be a projective plane, let S be a semifield of coordinates for π, and let A be a cube corresponding to S. Let S_1 be the pre-semifield described by $A^{(23)}$, and let S_2 be a semifield constructed from S_1 by isotopy. Then we define π^T, the transpose of π, to be the plane coordinatized by S_2.

Theorem 5.1.1. π^T is uniquely defined.

Proof: No matter which choice of S is made, the resulting A's will be equivalent, by Theorem 4.5.2. The 3-cube $A^{(23)}$ is nonsingular, by Theorem 4.3.1. If $A \equiv B$, then $A^{(23)} \equiv B^{(23)}$, because of Theorem 4.2.3. Therefore $A^{(23)}$ is uniquely determined up to equivalence. Thus the plane π^T is uniquely determined. ▯

Corollary 5.1.2. $(\pi^T)^T = \pi$. ▯

5.2. Dual defined. The dual π^D of a projective plane coordinatized by a semifield is well known to be determined by constructing the anti-isomorphic semifield, i.e., replacing ab by ba in the semifield. (This is easy to see by considering the homogeneous coordinate notation.) This definition can be phrased in the same way as the definition for π^T, by using $A^{(12)}$ rather than $A^{(23)}$. We may write $\pi^D = \pi^{(12)}$, $\pi^T = \pi^{(23)}$.

Theorem 5.2.1. The operations of dual and transpose generate a series of at most six planes, according to the following scheme:

$$\begin{tikzcd}
 & \pi \arrow[dl, leftrightarrow, "T"'] \arrow[dr, leftrightarrow, "D"] & \\
\pi^{(23)} \arrow[d, leftrightarrow, "D"'] & & \pi^{(12)} \arrow[d, leftrightarrow, "T"] \\
\pi^{(123)} \arrow[dr, leftrightarrow, "T"] & & \pi^{(132)} \arrow[dl, leftrightarrow, "D"'] \\
 & \pi^{(13)} &
\end{tikzcd}$$

Proof: This theorem is clear from the manner in which dual and transpose have been defined. ‖

Corollary 5.2.2. If $\pi = \pi^D \neq \pi^T$ *there exists a third plane* $\pi^{TD} \neq \pi$, $\pi^{TD} \neq \pi^T$.

Proof: If $\pi^{TD} = \pi$ then $\pi^T = \pi^D = \pi$. If $\pi^{TD} = \pi^T$, then $\pi^{TD} = \pi^{DTD} = \pi^{TDT} = \pi$. ‖

5.3. Collineations.

Theorem 5.3.1. The collineation group of π^T *has the same order as the collineation group of* π.

Proof: Refer to Theorem 3.4.4. The groups of translations and shears are ismorphic. The autotopism groups are also isomorphic, by Theorem 4.2.3, since

$$[F,G,H] \times A = A \quad \leftrightarrow \quad [F,H,G] \times A^{(23)} = A^{(23)}. \tag{5.1}$$

Therefore, the orders of the collineation groups are the same. The interrelations of autotopisms, translations, and shears are, however, different as equations (3.16) show. ‖

5.4. Examples.

The concept of transpose can be generalized to apply to Veblen-Wedderburn systems, but that will be omitted here.

The transpose of a Desarguesian plane is Desarguesian, for all matrices of left multiplication are powers of the same matrix, and this property remains under transposition.

The semifield planes of order 32 have been completely tabulated by R. J. Walker [4]. Besides the Desarguesian plane, there are five others. These calculations were independently checked by the present author and some interesting relationships were found.

Plane P(1) has the 3-cube representation

$$\begin{array}{ccccc} 10000 & 01000 & 00100 & 00010 & 00001 \\ 01000 & 00100 & 00010 & 00001 & 10100 \\ 00100 & 00010 & 01001 & 10100 & 00101 \\ 00010 & 00001 & 11010 & 11110 & 10111 \\ 00001 & 10010 & 11011 & 10000 & 01110 \end{array} . \tag{5.2}$$

No way to construct this plane, except by trial and error, is known. The solution given of all the representatives of the plane has, in some sense, a relatively high degree of symmetry. This plane was discovered in December, 1961, by a program written for a Burroughs 220 digital computer.

There are no autotopisms except the identity; hence, the collineation group of P(1) consists entirely of translations and shears. This is the only known semifield plane (except for its dual) with this property. A further consequence is that there are 31^2 distinct semifields, isotopic but not isomorphic (see equation (3.13)).

Plane P(2) is the dual and also the transpose of P(1), and it has the same properties.

Plane P(3) is constructed by the method of Section VIII. It has the following 3-cube representation:

$$\begin{array}{ccccc} 10000 & 01000 & 00100 & 00010 & 00001 \\ 01000 & 00100 & 00010 & 00001 & 11000 \\ 00100 & 00010 & 11001 & 01010 & 00111 \\ 00010 & 00001 & 01010 & 11011 & 11100 \\ 00001 & 11000 & 00111 & 11100 & 10111 \end{array} \tag{5.3}$$

This particular representative has five automorphisms and no other autotopisms; there are 192 other distinct systems isotopic to this one, and none of these have any automorphisms. This agrees with formula 3.13:

$$31^2 = (5/5) + 192\cdot(5/1).$$

Plane P(3) is self-dual since its semifield is commutative, but it is not self-transpose. Therefore, (Corollary 5.2.2) there are two other planes, $P(4) = P(3)^T$, and $P(5) = P(4)^D$. These planes also contain one system with automorphisms and 192 other systems without.

It has been shown, by exhaustive enumeration on a computer, that these five planes constitute the entire set of proper semifield planes of order 32.

VI. SOME KNOWN SEMIFIELDS.

In this section, we discuss many of the known semifields and consider what the possible orders of semifields are. Throughout this section the letter p will denote a prime number.

6.1. Orders which are excluded. We have seen that semifields must have p^n elements. For each p and n there is a unique field with p^n elements. What can be said about proper semifields?

If the order is p, a semifield must be the field $GF(p)$. Furthermore, if the order is p^2, it is well known that a semifield must be the field $GF(p^2)$: Let $\{1,x\}$ be a basis for the semifield; the multiplication is determined by the definition of $x^2 = ax + b$. But the polynomial $x^2 - ax - b$ has no roots in $GF(p)$, else we would have $(x-r)(x-s) = x^2 - ax - b = 0$ contradicting axiom A2. Thus, $x^2 - ax - b$ is irreducible, and the multiplication is that of $GF(p^2)$.

If the order is 8, one easily verifies that we have only the field $GF(8)$. This must be true since there is a unique projective plane of order 8, but we can verify this particular case directly: Let $\{1,x,y\}$ be a basis, and let L be the matrix A^{2**} for left multiplication by x. If the characteristic equation of L is

$\lambda^3 + a\lambda^2 + b\lambda + c = 0$, L satisfies this equation, so we find that $x(x^2) + ax^2 + bx + c = 0$. This polynomial must have no linear factors; hence, it is irreducible and takes one of the two forms $x(x^2)+x+1 = 0$ or $x(x^2)+x^2+1 = 0$. Replacing x by $x+1$, if necessary, we can assume $x(x^2) + x + 1 = 0$. In particular, $\{1,x,x^2\}$ is a basis.

The remainder of the proof considers various nontrivial possibilities and will only be sketched:

I. If $x^2x = x^2+1$, $x^2x^2 = x$, then $(x^2+x)^2 = 0$.

II. If $x^2x = x^2+1$, $x^2x^2 = x^2+x$, then $(x^2+1)(x^2+x+1) = 0$.

III. If $x^2x = x+1$, $x^2x^2 = 1$, then $(x^2+1)^2 = 0$.

IV. If $x^2x = x+1$, $x^2x^2 = x$, then $(x^2+x)(x^2+x+1) = 0$.

V. If $x^2x = x+1$, $x^2x^2 = x^2+1$, then $(x^2+1)(x^2+x) = 0$

VI. If $x^2x = x+1$, $x^2x^2 = x^2+x$, then we have the field $GF(8)$.

We have proved the following theorem:

Theorem 6.1. A proper semifield has order p^n, where $n \geq 3$ and $p^n \geq 16$. ∎

We will show in Section VIII that these easily obtained, necessary conditions on the order are actually sufficient.

6.2. Semifields of order 16. The semifields of order 16 have been tabulated in [5], and these have been independently checked on a computer. There are 23 nonisomorphic, proper semifields of order 16. These are all isotopic to either system V or to system W of Section 2.2; consequently, two projective planes are formed.

The first plane, consisting of those semifields isotopic to V, contains 18 distinct semifields. There is one (namely V) with 6 automorphisms, another with 3 automorphisms, 8 with 2 automorphisms, and 8 with only the identity automorphism. Hence, there must be 18 autotopisms in agreement with formula (3.13):

$$15^2 = 18(\frac{1}{6} + \frac{1}{3} + \frac{8}{2} + \frac{8}{1}). \tag{6.1}$$

The second plane has only five distinct semifields. One of these has 4 automorphisms, one (W) has 3, and the other three have 2 automorphisms. Thus, by formula 3.13, there are 108 autotopisms, and

$$15^2 = 108(\frac{1}{4} + \frac{1}{3} + \frac{3}{2}). \tag{6.2}$$

Since the number of autotopisms is different, we can conclude that each plane is self-dual and self-transpose.

6.3. Early work of Dickson. The study of semifields apparently was originated by L. E. Dickson in 1905 [6]. In two early papers on the subject, he considered the construction of all possible semifields of order p^3, and of all possible commutative semifields of order p^4, where p is odd.

Perhaps the simplest way to construct proper semifields is analogous to our construction of V and W in Section II; we start with a field $GF(p^m)$ and construct a semifield of order p^{2m}, having elements $a + \lambda b; \quad a,b \in GF(p^m)$.

One must merely define

$$(a + \lambda b)(c + \lambda d) = f(a,b,c,d) + \lambda g(a,b,c,d) \qquad (6.3)$$

where f,g are linear in all four variables, and where

$f(a,b,c,d) = g(a,b,c,d) = 0$ implies $a = b = 0$ or $c = d = 0$.

There are many, many ways to do this, seemingly unrelated; a quite general class of these constructions is investigated in Section VII.

Dickson [7] gave a particularly simple construction of a commutative semifield of this type, for p odd:

$$(a + \lambda b)(c + \lambda d) = (ac + b^\sigma d^\sigma f) + \lambda(ad+bc) \qquad (6.4)$$

where σ is an automorphism and where f is a nonsquare of the field $GF(p^m)$. The condition that f must be a nonsquare element is clear, for if $f = a^2$ choose b such that $b^\sigma = a^{-1}$. Then $(1 + \lambda b)(1 - \lambda b) = 0$. It is easy to verify that the condition is also sufficient (see Section VII).

The complex numbers are a particular case of the system (6.4), although f being a nonsquare is not always sufficient in the infinite case. The system (6.4) is associative if and only if $\sigma = I$.

6.4. Twisted fields. The following construction is due to A. A. Albert [2,8,9]. Define a new multiplication on the elements of $GF(p^n)$ by

$$x \circ y = xy^q - c\, x^q y, \tag{6.5}$$

where $q = p^m$, $1 \leqq m < n$, and where $c \neq a^{q-1}$ for $a \varepsilon GF(p^n)$. Then we obtain a pre-semifield, since $x^q + y^q = (x + y)^q$, and since

$$x \circ y = 0 \text{ implies } x = 0 \text{ or } y = 0 \text{ or } c = (y/x)^{q-1}.$$

Now pass to a semifield by using (4.19); the result is called a twisted field.

The construction can be carried out only if c exists subject to the required conditions. This will be the case when $(q-1, p^n-1) > 1$, i.e., when $q-1$ and p^n-1 have a common factor, since multiplication is cyclic of order p^n-1. But if p is odd, there is always the common factor $p - 1$; if $p = 2$ we have $(2^m-1, 2^n-1) = 2^{(m,n)}-1$, so we need $(m,n) > 1$. If $n = mk$, where $k > 2$, it has been shown that the semifield constructed is nonassociative.

Twisted fields exist for nearly all orders not excluded by Theorem 6.1.1. The missing orders are 2^4, and 2^p, where p is a prime greater than 3.

6.5. Sandler's construction. An interesting class of semifields has been constructed by R. Sandler [10]. There are p^{nm^2} elements in these semifields, where m is greater than 1.

For Sandler's construction, let $q = p^n$; the elements of S are

$$a_0 + \lambda a_1 + \cdots + \lambda^{m-1} a_{m-1}, \quad a_i \in GF(q^m). \tag{6.6}$$

Multiplication is defined as follows:

$$(\lambda^i x)(\lambda^j y) = \lambda^{i+j} x^{q^j} y, \quad 0 \leqq i < m, \quad 0 \leqq j < \infty.$$

$$\lambda^m = \delta \; , \tag{6.7}$$

with the convention that λ^k denotes left powers of λ, i.e., $\lambda^{k+1} = \lambda\lambda^k$. If δ is chosen such that it satisfies no polynomial of degree less than m over $GF(q)$, this gives a semifield.

For example, let us construct such a system S of order 2^9. The elements are

$$a + \lambda b + \lambda^2 c, \quad a,b,c \in GF(8).$$

Multiplication is defined by the rule

$$\begin{aligned}(a + \lambda b + \lambda^2 c)(d + \lambda e + \lambda^2 f) = \quad & ad + \lambda a^2 e + \lambda^2 a^4 f \\ & + \delta b^4 f + \lambda bd + \lambda^2 b^2 e \\ & + \delta c^2 e + \lambda \delta c^4 f + \lambda^2 cd.\end{aligned} \tag{6.8}$$

Note the similarity between this and the definition of the system W in Section 2.2. This can be written in matrix form, since S is a right vector space over $GF(8)$; the matrix of left multiplication by $(a + \lambda b + \lambda^2 c)$ is then

$$L = \begin{bmatrix} a & \delta c^2 & \delta b^4 \\ b & a^2 & \delta c^4 \\ c & b^2 & a^4 \end{bmatrix}. \tag{6.9}$$

The determinant of L is $a^7 + \delta b^7 + \delta^2 c^7 - \delta(a^2b^4c + ab^2c^4 + a^4bc^2) = r + s\delta + t\delta^2$. Since $r^2 = r$, $s^2 = s$, $t^2 = t$ this is a polynomial of degree 2 or less over $GF(2)$, and cannot be zero by hypothesis unless all of the coefficients vanish; but this would require $a = b = c = 0$.

VII. WEAK NUCLEI AND QUADRATIC EXTENSIONS.

The concept of nucleus (Section 2.6) is generalized here, and this leads to an important class of semifields of the form (6.3).

7.1. Definition of weak nucleus.

Let F be a field contained in a semifield S. We say F is a weak nucleus for S if $(ab)c = a(bc)$ whenever any two of a,b,c are in F. S will be a vector space over F, but left and right multiplication will not in general be linear transformations over F.

An example of such a system is the semifield V of Section 2.2. Here F is contained in neither the left, middle, or right nucleus of V. There are, on the other hand, semifields of order 16 whose left nucleus is not a weak nucleus. In other words, the statement, "F is a weak nucleus" does not imply and is not implied by the statement "F is a left nucleus." However, if F is contained in any two of the nuclei N_ℓ, N_m, or N_r, F must necessarily be a weak nucleus.

7.2. Quadratic extensions.

Theorem 7.2.1. *Let* F *be a weak nucleus for* S, *and let* S *have dimension 2 over* F. *Then the elements of* S *have the form*

$$a + \lambda b, \qquad a, b \in F; \tag{7.1}$$

and λ *can be chosen such that*

$$a\lambda = \lambda a^{\sigma} \tag{7.2}$$

for all $a \varepsilon F$, *where* σ *is an automorphism of* F.

Proof: Let x be a primitive element of the field F, and let

$$x\lambda = a + \lambda b. \tag{7.3}$$

Case I. $b = x$. Then we can prove by induction that

$$x^{n}\lambda = nax^{n-1} + \lambda x^{n}; \tag{7.4}$$

$$\text{for} \quad x^{n+1}\lambda = (xx^{n})\lambda = x(x^{n}\lambda) = x(nax^{n-1} + \lambda x^{n}) = nax^{n} + x(\lambda x^{n})$$

$$= nax^{n} + (x\lambda)x^{n} = nax^{n} + (a + \lambda x)x^{n} = (n+1)ax^{n} + \lambda x^{n+1}.$$

Notice that the associative law was applied here only when permissible. If F has p^{m} elements, we can put $n = p^{m}$, and since $p^{m}a = 0$ and $x^{p^{m}} = x$, we have by (7.4)

$$x\lambda = \lambda x. \tag{7.5}$$

Case II. $b \neq x$. Then we may replace λ by $\lambda' = \lambda + a(b-x)^{-1}$, and we find $x\lambda' = \lambda' b$. We easily prove by induction that $x^{n}\lambda' = \lambda' b^{n}$. Since x is a primitive element, b is a power of x, say x^{σ}. This shows that $y\lambda' = \lambda' y^{\sigma}$ for all nonzero y, and in fact

$$(yz)\lambda' = \lambda'(y^{\sigma}z^{\sigma}), \tag{7.6}$$

for all nonzero y,z since y and z are powers of x. Furthermore,

$$(y + z)\lambda' = \lambda'(y^\sigma + z^\sigma), \tag{7.7}$$

by the distributive laws. These equations hold trivially if y or z is zero; therefore, we have shown that $(yz)^\sigma = y^\sigma z^\sigma$, $(y+z)^\sigma = y^\sigma + z^\sigma$. This means σ is an automorphism, i.e., $\sigma = p^k$ for some k. ▮

7.3. Construction of semifields. The rule for multiplication in a quadratic extension of a weak nucleus must be

$$(a + \lambda b)(c + \lambda d) = ac + \lambda(a^\sigma d + bc) + (\lambda b)(\lambda d), \tag{7.8}$$

according to the preceding theorem. But the fact that F is a weak nucleus tells us nothing about what the product $(\lambda b)(\lambda d)$ must be; such products never occur unless two elements not in F are multiplied together. This product can be defined as $f(b,d) + \lambda g(b,d)$ where f and g are bilinear functions, as long as no zero divisors are introduced.

Although many choices might be made for f and g, we will make the assumption here that they have a certain simple form.

Let
$$(\lambda b)(\lambda d) = b^\alpha d^\beta f + \lambda(b^\gamma d^\delta g) \tag{7.9}$$

where f,g are now elements of F, and where $\alpha,\beta,\gamma,\delta$ are automorphisms of F. Under what conditions can we conclude that no zero divisors have been introduced?

Suppose $(a+\lambda b) \neq 0$, $(c + \lambda d) \neq 0$, but $(a + \lambda b)(c + \lambda d) = 0$. Then, in the first place, we have

$$ac + b^{\alpha}d^{\beta}f = 0;$$

this implies there is an element $x \neq 0$ in F such that

$$a = xd^{\beta}, \quad c = -x^{-1}b^{\alpha}f. \tag{7.10}$$

The other condition is that

$$a^{\sigma}d + bc + b^{\gamma}d^{\delta}g = 0, \text{ i.e.,}$$

$$x^{\sigma}d^{\beta\sigma+1} - b^{\alpha+1}x^{-1}f + b^{\gamma}d^{\delta}g = 0 \tag{7.11}$$

Case I. $g = 0$. If $g = 0$, we have $x^{\sigma}d^{\beta\sigma+1} = b^{\alpha+1}x^{-1}f$. Here b and d must be nonzero, and $f = x^{\sigma+1}d^{\beta\sigma+1}b^{-1-\alpha}$. Therefore, if p is an odd prime, $\sigma+1$, $\beta\sigma+1$, and $-1-\alpha$ are even and we could conclude that f is a square; so choose f to be a nonsquare.

Other a priori choices of α,β and f can be made, in which f is a square, as we will see in our discussion of Case II. For particular fields, special choices may be possible.

Case II. $g \neq 0$. In this case, we will try to choose $\alpha,\beta,\gamma,\delta$ so that (7.11) is a nonzero constant times a polynomial in a single variable. This can be done in essentially one way by writing it in the form

$$x^{-1}b^{\alpha+1}(x^{\sigma+1}b^{-1-\alpha}d^{\beta\sigma+1} + xb^{\gamma-1-\alpha}d^{\delta}g - f) = 0,$$

and we require that

$$x^{\sigma+1}b^{-1-\alpha}d^{\beta\sigma+1} = (xb^{\gamma-1-\alpha}d^{\delta})^{\sigma+1}. \tag{7.12}$$

Theorem 7.3.1.

$$p^a + p^b \equiv p^c + p^d \pmod{p^m-1}$$

if and only if $a \equiv c,\ b \equiv d$ or $a \equiv d,\ b \equiv c \pmod{m}$.

Proof: Since $p^m \equiv 1 \pmod{p^m-1}$, we may assume $0 \leqq a,b,c,d < m$. Then we have $2 \leqq p^a + p^b \leqq 2p^{m-1} \leqq p^m$, and $2 \leqq p^c + p^d \leqq 2p^{m-1} \leqq p^m$, so we may conclude that

$$p^a + p^b = p^c + p^d.$$

We may now divide by p if necessary, until one of a,b,c,d is zero, say a. Then it is clear that either c or d must be zero, and the theorem follows. ▌

Now suppose $F = GF(p^m)$, and let $\tau = \sigma^{-1}$. Then (7.12) can be written

$$\begin{aligned} \beta\sigma + 1 &\equiv \delta\sigma + \delta \pmod{p^m-1} \\ \gamma\sigma + \gamma &\equiv \alpha\sigma + \sigma \pmod{p^m-1} \end{aligned} \tag{7.13}$$

By Theorem 7.3.1, these two congruences have four solutions:

$$(\beta,\delta) = (\tau^2,\tau) \text{ or } (1,1)\ ;\ (\alpha,\gamma) = (\tau,1) \text{ or } (\sigma,\sigma). \tag{7.14}$$

Using any of these four solutions, we can conclude that there are no zero divisors if and only if the polynomial

$$y^{\sigma+1} + gy - f = 0 \tag{7.15}$$

has no solutions in F.

There are in general many such polynomials, as long as F is not a prime field. For example, if p is odd, we may take $g = 0$, $\sigma = p$ and f any element not of the form y^{p+1}. If $p = 2$, we can take $g = 1$ and $\sigma = 2$, and f any element not of the form $y^3 + y$. (Such an element must exist, since $0^3 + 0 = 1^3 + 1$.)

7.4. <u>The four types</u>. We obtain many proper semifields, which are quadratic over a weak nucleus, by the construction in the previous section. Those of Case I have the multiplication rule

$$(a + \lambda b)(c + \lambda d) = (ac + b^{\alpha}d^{\beta}f) + \lambda(a^{\sigma}d + bc) \tag{7.16}$$

where p is odd and f is a nonsquare; α, β, and σ are arbitrary automorphisms, but not all the identity. Dickson's construction (see Section 6.3) is a special case, indeed the only commutative semifield of this type.

Other semifields, those of Case II, are constructed by finding an automorphism $\sigma \neq I$ and elements $f, g \in F$ such that

$$y^{\sigma+1} + gy - f \neq 0, \quad \text{for } y \in F.$$

Then four types of semifields are produced, all for the same σ, f, g and for $\tau = \sigma^{-1}$:

$$(a + \lambda b)(c + \lambda d) =$$

$$\begin{array}{rlcl} \text{I.} & (ac + b^{\sigma}d^{\tau^2}f) & + & \lambda(bc + a^{\sigma}d + b^{\sigma}d^{\tau}g) \\ \text{II.} & (ac + b^{\sigma}d\ \ f) & + & \lambda(bc + a^{\sigma}d + b^{\sigma}d\ \ g) \\ \text{III.} & (ac + b^{\tau}d^{\tau^2}f) & + & \lambda(bc + a^{\sigma}d + b\ d^{\tau}g) \\ \text{IV.} & (ac + b^{\tau}d\ \ f) & + & \lambda(bc + a^{\sigma}d + b\ d\ g) \end{array} \tag{7.17}$$

For example, the system V of Section 2.2 is of type I. The system W of that section is actually of all four types. The systems of type II were discovered by Hughes and Kleinfeld [11]. Autotopisms of systems of type II are discussed in [10,12,13].

Theorem 7.4.1. *Let* S *be a proper semifield, quadratic over a finite field* F. *Let* N_ℓ, N_m, N_r *be the nuclei of* S. *Then*

(a) $F = N_r = N_m$ *if and only if* S *is of type II*.

(b) $F = N_\ell = N_m$ *if and only if* S *is of type III*.

(c) $F = N_\ell = N_r$ *if and only if* S *is of type IV*.

Proof: The "if" parts of this theorem can be obtained by direct calculation, and will be omitted. The "only if" parts follow since, first, F is a weak nucleus if it equals any two of the nuclei, so Theorem 7.2.1 holds. We need then merely calculate $(\lambda b)(\lambda d)$ under the assumptions (a), (b), (c) given.

Let $\lambda^2 = f + \lambda g$. Then

(a)
$$(\lambda b)(\lambda d) = ((\lambda b)\lambda)d = (\lambda(b\lambda))d = (\lambda(\lambda b^\sigma))d = \lambda((\lambda b^\sigma)d)$$
$$= \lambda(\lambda(b^\sigma d)) = \lambda^2(b^\sigma d) = b^\sigma df + \lambda b^\sigma dg.$$

(b)
$$(\lambda b)(\lambda d) = (b^\tau\lambda)(d^\tau\lambda) = b^\tau(\lambda(d^\tau\lambda)) = b^\tau((\lambda d^\tau)\lambda) = b^\tau((d^{\tau^2}\lambda)\lambda)$$
$$= b^\tau(d^{\tau^2}\lambda^2) = (b^\tau d^{\tau^2})\lambda^2 = b^\tau d^{\tau^2}(f + \lambda g)$$
$$= b^\tau d^{\tau^2} f + \lambda b d^\tau g.$$

(c)
$$(\lambda b)(\lambda d) = (b^\tau\lambda)(\lambda d) = b^\tau(\lambda(\lambda d)) = b^\tau(\lambda^2 d)$$
$$= b^\tau((f + \lambda g)d) = b^\tau df + \lambda bdg.$$

The proof is completed by observing that either of the three types implies the existence of the polynomial (7.15) which has no solutions in F. ∎

No characterisation of type I is known.

Corollary 7.4.2. *Under the hypotheses of Theorem 7.4.1,* $F = N_\ell = N_m = N_r$ *if and only if* S *is of all four types I, II, III, IV where* $\sigma^2 = 1$ *and* $g = 0$.

Proof: By the theorem, we must have S of types II, III, IV at least. These types are the same if and only if $\sigma^2 = 1$ and $g = 0$, since $\sigma \neq 1$. ∎

This corollary and part (a) of Theorem 7.4.1 were discovered by Hughes and Kleinfeld in their original paper [11].

VIII. BINARY SEMIFIELDS.

Constructions given in this section fill the gap in the existence theorems for semifields and non-Desarguesian planes of all possible orders p^n. Proper semifields of all orders 2^n, where $n \neq 3$ and n is not a power of 2, are constructed.

8.1. The binary pre-semifields. Let $K = GF(2^{mn})$ where n is odd, $n > 1$; let K_o be the subfield $GF(2^m)$. Considering K as a vector space over K_o, let f be any nonzero linear functional from K to K_o:

$$f(xa + yb) = xf(a) + yf(b), \tag{8.1}$$

for all $a,b \in K$, and all $x,y \in K_o$.

Define a new multiplication in K as follows:

$$a \circ b = ab + [f(a)b + f(b)a]^2. \tag{8.2}$$

This product clearly is linear in both variables, so it satisfies both distributive laws. We will show that there are no zero divisors: Suppose $a \circ b = 0$, and $a,b \neq 0$; then let $x = ab^{-1}$. This implies

$$x + f(a)^2 + f(b)^2x^2 = 0.$$

We have a quadratic equation with coefficients in K_o; but since the degree of K/K_o is odd, this equation must be reducible. Therefore, $x \varepsilon K_o$. But then $a = xb$ implies

$$a \circ b = ab + [f(xb)b + f(b)xb]^2 = ab \neq 0.$$

<u>8.2. The binary semifields.</u> A semifield can be obtained from this pre-semifield in any of the standard ways described in Section 4.5; let us call any such semifield a <u>binary semifield</u>. We will give a particular construction in detail here. It is necessary to do this, to make sure the resulting semifield is not a field.

Let $\{1,x,x^2,\ldots,x^{n-1}\}$ be a basis of K over K_o; set

$$f(1) = f(x) = \cdots = f(x^{n-2}) = 0, \quad f(x^{n-1}) = 1. \tag{8.3}$$

With this definition, we find that, for $y \varepsilon K_o$,

$$1 \circ y = y,\ 1 \circ yx = yx,\ \ldots,\ 1 \circ yx^{n-2} = yx^{n-2},$$

$$1 \circ yx^{n-1} = yx^{n-1} + y^2;$$

and therefore $\quad 1 \circ (1 \circ a) = a, \quad$ all $\ a \varepsilon K.$ $\hfill (8.4)$

Define the multiplication $*$ by the rule

$$a * b = (1 \circ a) \circ (1 \circ b). \tag{8.5}$$

Then 1 is an identity for $a * b$, and we have a semifield $(S,+,*)$.

Notice that we have now defined three different multiplications on the elements of K: ab, $a \circ b$, and $a * b$. It is important to keep this distinction in mind, since all three multiplications are used simultaneously in many proofs. The powers of an element a^2, a^3, etc., will always refer to the multiplication of the field.

Theorem 8.2.1. S is a proper semifield of order 2^{mn}, if n is odd, and if $mn > 3$.

Proof: If $n > 3$, let $k = (n-1)/2$; then $k > 1$, and

$$x * (x^k * x^k) = x * x^{n-1} \neq x^n = (x * x^k) * x^k.$$

Thus, multiplication is not associative in this case.

If $n = 3$, let y be an element of K_0; we have

$$(x * x) * yx = x^2 * yx = (x^2 + 1) \circ yx = yx(x^2+1) + y^2x^2.$$

$$x * (x * yx) = x * yx^2 = x \circ (yx^2+y^2) = (yx^2+y^2)x + y^2x^2.$$

Thus, multiplication is still not associative, unless $y^2 = y$. We can always choose $y \neq y^2$ unless $K = GF(8)$, in which case we have shown before that no proper semifield exists. ▮

Corollary 8.2.2. (Converse of Theorem 6.1) Proper semifields exist for all orders p^n, where $n \geq 3$ and $p^n \geq 16$. ▮

Corollary 8.2.3. Non-Desarguesian projective planes exist for all orders p^n, provided $n \geq 2$ and $p^n \geq 9$.

Proof: Hall systems (Veblen-Wedderburn systems) exist for the remaining orders p^2, p odd. ▮

8.3. Isotopes. The definition in 8.1 involves an arbitrary linear functional f. In this section, however, we show that all of the binary semifields are isotopic, so only one plane is determined.

Theorem 8.3.1. *If f,g are nonzero linear functionals from K to K_0, there exists an element $z \in K$ such that $f(az) = g(a)$, for all a in K.*

Proof: A simple counting argument applies. Let $\{x_1, x_2, \ldots x_n\}$ be a basis of K over K_0. A linear functional is completely specified by giving $f(x_i)$, $1 \leq i \leq n$, and these choices are independent; hence there are $2^{nm} - 1$ nonzero linear functionals.

Suppose f is a nonzero linear functional; then $f(xz)$ is also a nonzero linear functional, if $z \neq 0$. There are $2^{nm} - 1$ choices for z, so we need only show that no two of these give the same function.

But if $f(az_1) = f(az_2)$ for all a, we have $f(a(z_1 - z_2)) = 0$ for all a, hence $z_1 - z_2 = 0$. ▌

Theorem 8.3.2. *All binary semifields, for given K and K_0, are isotopic.*

Proof: We need only show that any two of the *pre*-semifields are isotopic. Suppose we have

$$a \circ b = ab + [f(a)b + f(b)a]^2$$

$$a \cdot b = ab + [g(a)b + g(b)a]^2.$$

By Theorem 6.3.1, we find $z \varepsilon K$ such that $f(az) = g(a)$ for all a. Then we have

$$az \circ bz = abz^2 + [g(a)bz + g(b)az]^2 = (a \cdot b)z^2. \quad \|$$

8.4. Automorphisms. The binary semifield exhibited for the plane $P(3)$ in Section 5.4 has 5 automorphisms. This property holds in general: we will show there is a binary semifield with n automorphisms.

In this section we let $q = 2^m$.

Theorem 8.4.1. Let f *be such that* $f(a) = x$ *whenever*

$$a = x + b + b^q, \quad x \varepsilon K_o, \quad b \varepsilon K. \tag{8.6}$$

Then f *is a linear functional from* K *to* K_o.

Proof: First we show that f is well-defined. Suppose $x + b + b^q = y + c + c^q$ for $x \neq y$; then $(b+c)^q = (b+c) + x+y$, i.e., $a^q = a + z$, for some $a \varepsilon K$, $z \varepsilon K_o$. Applying this rule again, we find

$$a^{q^2} = a^q + z^q = a^q + z = a.$$

Since n is odd, we find $a^{q^{n+1}} = a$. But $a^{q^{n+1}} = a^q$, hence $z = 0$. Thus f is well-defined.

Furthermore,

$$f(x+b+b^q + y+c+c^q) = f(x+y + (b+c) + (b+c)^q) = x + y$$
$$= f(x+b+b^q) + f(y+c+c^q).$$

$$f(y(x+b+b^q)) = f(yx + yb + (yb)^q) = yx$$
$$= yf(x+b+b^q).$$

This shows that f is a linear functional. $\|$

Theorem 8.4.2. Let f *be defined as in Theorem 8.4.1, and let the product* $a \circ b$ *be defined by equation (8.2). The multiplication* $*$ *defined by*

$$(1 \circ a) * (1 \circ b) = a \circ b \tag{8.7}$$

has the automorphism $a \to a^q$, *hence there are at least* n *automorphisms of the binary semifield.*

Proof: We show first that $a \to a^q$ is an automorphism of the circle product $a \circ b$. (By a^q we mean the multiplication of the field K.)

$$\begin{aligned}(a \circ b)^q &= (ab + [f(a)b + f(b)a]^2)^q \\ &= a^q b^q + [f(a)b^q + f(b)a^q]^2 \\ &= a^q b^q + [f(a^q)b^q + f(b^q)a^q]^2 = a^q \circ b^q,\end{aligned}$$

since $f(a) = f(a^q)$ by definition.

The automorphism carries over to the star product, since

$$\begin{aligned}((1\circ a) * (1\circ b))^q &= (a\circ b)^q = a^q \circ b^q \\ &= (1\circ a^q) * (1\circ b^q) \\ &= (1^q\circ a^q) * (1^q\circ b^q) = (1\circ a)^q * (1\circ b)^q. \quad \blacksquare\end{aligned}$$

8.5. Autotopisms. Let us investigate the autotopism group of a binary semifield; it suffices to consider autotopisms of the presemifields. The following theorem is an important step in this direction:

Theorem 8.5.1. If $[F,G,H]$ is an autotopism of a binary pre-semifield, we have $F = Gz$, for $z \neq 0$ in K_o; i.e.,

$$aF = z(aG), \text{ for all } a \in K. \tag{8.8}$$

Proof: We have

$$aF \circ bG = (a \circ b)H = (b \circ a)H = bF \circ aG, \text{ for all } a,b.$$

Therefore

$$(aF)(bG) + [f(aF)bG + f(bG)aF]^2 = (aG)(bF) + [f(aG)bF + f(bF)aG]^2. \tag{8.9}$$

Let $V_1 = \{a \mid f(aF) = 0\}$, $V_2 = \{a \mid f(aG) = 0\}$.

V_1 and V_2 are vector spaces of dimension $m(n-1)$ over $GF(2)$, since the kernel of $f: K \to K_o$ must have dimension $n-1$ over K_o.

Let $V_3 = V_1 \cap V_2$. Then

$$\dim V_3 = \dim V_1 + \dim V_2 - \dim V_1 \cup V_2 \geqq 2m(n-1)-mn = m(n-2) \geqq 2.$$

Therefore V_3 contains at least two nonzero elements.

We apply formula (8.9), to find,

$$(aF)(bG) = (aG)(bF), \text{ if } f(aF) = f(aG) = f(bF) = f(bG) = 0;$$

i.e.,
$$\frac{aF}{aG} = \frac{bF}{bG} = z, \text{ for all } a,b \in V_3 - \{0\}. \tag{8.10}$$

Now, let $a \in V_3 - \{0\}$ and let b be arbitrary; we have

$$z(aG)(bG) + [zf(bG)(aG)]^2 = (aG)(bF) + [f(bF)(aG)]^2,$$

i.e.,
$$z(bG) + bF = (aG)[zf(bG) + f(bF)]^2. \tag{8.11}$$

Let a run through two distinct elements a_1, a_2 of $V_3 - \{0\}$; the left hand side stays constant, hence we find

$$(a_1G - a_2G)[zf(bG) + f(bF)]^2 = 0.$$

But G is 1-1, so $a_1G - a_2G \neq 0$. This shows us that

$$zf(bG) + f(bF) = 0, \quad \text{all } b.$$

In particular, if $b \notin V_2$, we find $z \in K_o$.

Finally, equation (8.11) becomes

$$z(bG) + bF = 0, \quad \text{all } b,$$

and this is precisely equation (8.8). ▮

Remarks: Put $a = b$, we find $(a \circ a)H = (a^2)H = z(aG)^2$, so H is given in terms of G. Other relationships can now be obtained, using methods similar to those of the preceding theorem, but this investigation is incomplete at the present time.

References

[1] M. Hall, Jr. The Theory of Groups. New York, Macmillan, 1959. pp. 346-420.

[2] A. A. Albert, "Finite division algebras and finite planes," Proc. Sym. Appl. Math. X (1960) pp. 53-70.

[3] Günter Pickert. Projektiv Ebenen. Berlin, Springer, 1955. 343 pp.

[4] R. J. Walker, "Determination of division algebras with 32 elements," Proc. of Sym. Exp. Arith., to appear.

[5] Erwin Kleinfeld, "Techniques for enumerating Veblen-Wedderburn systems," J. ACM 7 (1960), pp. 330-337.

[6] L. E. Dickson, "Linear algebras in which division is always uniquely possible," Trans. AMS 7(1906), pp. 370-390, 514-527.

[7] L. E. Dickson, "Linear algebras with associativity not assumed," Duke Math. J. 1(1935), pp. 113-125.

[8] A. A. Albert, "On non-associative division algebras," Trans. AMS 72 (1952), pp. 296-309.

[9] A. A. Albert, "Finite non-commutative division algebras," Proc. AMS 9(1958), pp. 928-932.

[10] Reuben Sandler, "Autotopism groups of some finite non-associative algebras," Amer. J. Math. 84 (1962), pp. 239-264.

[11] D. R. Hughes and Erwin Kleinfeld, "Semi-nuclear extensions of Galois fields," Amer. J. Math. 82 (1960), pp. 389-392.

[12] D. R. Hughes, "Collineation groups of non-Desarguesian planes II," Amer. J. Math., 82(1960), pp. 113-119.

[13] A. A. Albert, "On the collineation groups of certain non-Desarguesian planes," Portugaliae Mathematica, 18(1959), pp. 207-224.

[14] D. E. Knuth, "Non-Desarguesian planes of order 2^{2m+1}," (Abstract) AMS Notices 9 (June 1962), p. 218.